03-289r2 Only If Reserved Proposal

Date: November 4, 2003

To: T10 Committee (SCSI)

From: Roger Cummings (VERITAS)

Subject: T10/03-289r2 Only If Reserved Proposal

Revision History

03-289r0 (August 27, 2003) First Revision
03-289r1 (November 4, 2003) Clarified the specific commands to which the OIR bit applies.
03-289r2 (November 7, 2003) Proposed that bit be included in SSC Device Configuration page rather than Control Mode page.

Related Documents

03-231r0 (July 2, 2003) Two persistent reservations problems - latest information. Feedback on earlier approaches and results from latest testing & investigation.
spc3r15 – SCSI Primary Commands – 3 revision 15
ssc2r09 - SCSI Stream Commands - 2 revision 9

Background

This proposal has arisen from recent research into persistent reservations, but it is a separate proposal which stands alone from that work. The feature proposed here will work with (Original) Reservations, or any type of Persistent Reservations, and will directly address some significant problems encountered by designers of backup and restore applications. Adoption of this proposal will allow simplification of existing code, and provide much-improved resistance against a particularly destructive class of data corruption errors that have been encountered in a number of situations in the field.

Overview

Reserves and Releases are presently used extensively in situations where sequential-access devices are employed. There are known to be situations where multiple reserves and releases are issued against the same device by applications, device drivers and other entities that coexist in a computer system.

These situations occur because there is no way to test that an (original) reservation is still in existence for an I_T nexus. Issuing a RESERVE command returns the same status if a reservation to the issuing I_T nexus existed before the command was received or not. Similarly, issuing a RELEASE command returns the same status if a reservation to the issuing I_T nexus existed before the command was received or not.

Designers of backup and restore applications are very concerned to avoid writing to unreserved serial access devices, because they can be subject to later and almost undetectable data corruption resulting from accesses from other Initiators (either servers or data copy engines). The combination of this concern and the lack of a way of testing a reservation results in the profligate use of reserves described above.

However the provision of a test facility would not by itself solve this problem, because by definition there will always be a time interval between the completion of the test and the issuance of a data transfer command. To avoid this “hole” what is needed is a way to ensure that a command will only be executed by the device server if a reservation exists for the I_T nexus under which the command is issued. If no such reservation exists, the command would not be performed, and the device server would return a new status to indicate that the device server is not reserved.

Proposal

The document proposes the definition of a new bit for the SSC-3 Device Configuration mode page, called Only If Reserved (OIR). The definition of this bit has been extended from the background given above to also cover persistent reservations. A new ASC/ASCQ is also requested (to be defined in SPC-3).

Suggested Changes

8.8.3 Device Configuration mode page

The Device Configuration mode page (see table 60) is used to specify the appropriate sequential-access device configuration.

Table 60 — Device Configuration mode page

Bit Byte	7	6	5	4	3	2	1	0
0	PS	Rsvd	PAGE CODE (10h)					
1	PAGE LENGTH (0Eh)							
2	Rsvd	Obsolete	CAF	ACTIVE FORMAT				
3	ACTIVE PARTITION							
4	WRITE OBJECT BUFFER FULL RATIO							
5	READ OBJECT BUFFER EMPTY RATIO							
6	(MSB)	WRITE DELAY TIME						
7								(LSB)
8	OBR	LOIS	RSMK	AVC	SOCF		ROBO	REW
9	GAP SIZE							
10	EOD DEFINED			EEG	SEW	SWP	BAML	BAM
11	(MSB)	OBJECT BUFFER SIZE AT EARLY WARNING						
12								
13								(LSB)
14	SELECT DATA COMPRESSION ALGORITHM							
15	Reserved		OIR	REWIND ON RESET		ASOCWP	PERSWP	PRMWP

The parameters savable (PS) bit is only used with the MODE SENSE command. This bit is reserved with the MODE SELECT command. A PS bit of one specifies the logical unit is capable of saving the mode page in a non-volatile vendor-specific location.

NOTE 42 The change active partition (CAP) bit (byte 2, bit 6 in the Device Configuration mode page) has been obsoleted. To change active partitions refer to the LOCATE command.

A change active format (CAF) bit of one specifies the active format is to be changed to the value specified in the ACTIVE FORMAT field. A CAF bit of zero specifies no active format change is specified. For some devices, the format may only be changed when the logical unit is at beginning-of-partition.

The ACTIVE FORMAT field specifies the recording format that is in use for the selected density code when reading or writing data on a logical unit. The value of the ACTIVE FORMAT field is vendor-specific.

The ACTIVE PARTITION field specifies the current logical partition number in use on the medium. This shall be a non-changeable field.

The WRITE OBJECT BUFFER FULL RATIO field, on WRITE commands, specifies to the device server how full the object buffer shall be before writing data to the medium. A value of zero specifies the value is not specified.

The READ OBJECT BUFFER EMPTY RATIO field, on READ commands, specifies to the device server how empty the object buffer shall be before retrieving additional data from the medium. A value of zero specifies the value is not specified.

The WRITE DELAY TIME field specifies the maximum time, in 100 ms increments, that the device server should wait before any buffered data that is to be written, is forced to the medium after the last buffered WRITE command that did not cause the object buffer to exceed the write object buffer full ratio. A value of zero specifies the device server shall never force buffered data to the medium under these conditions.

An object buffer recovery (OBR) bit of one specifies the logical unit supports object buffer recovery using the RECOVER BUFFERED DATA command. An OBR bit of zero specifies the logical unit does not support object buffer recovery. Most device servers consider this bit to be not changeable.

A logical object identifiers supported (LOIS) bit of zero specifies logical object identifiers are not supported in the format written on the medium. A LOIS bit of one specifies the format on the medium has recorded information about the logical object identifiers relative to a partition. Most device servers consider this bit to be not changeable.

A report setmarks (RSMK) bit of one specifies the device and recording format supports setmarks. If the RSMK bit is set to one, the device server shall recognize and report setmarks during appropriate read or space operations. An RSMK bit of zero specifies the device or recording format does not support setmarks. This shall be a non-changeable bit.

The automatic velocity control (AVC) bit of one, specifies the device shall select the speed (if the device supports more than one speed) based on the data transfer rate that should optimize streaming activity and minimize medium repositioning. An AVC bit of zero specifies the speed chosen shall be defined by the SPEED field in the mode parameter header.

A stop on consecutive filemarks (SOCF) field of 00b specifies the device server shall pre-read data from the medium to the limits of the object buffer capacity without regard for filemarks. Values 01b, 10b, and 11b specify that the device server shall terminate the pre-read operation if one, two, or three consecutive filemarks are detected, respectively. If the RSMK bit is one, the device server shall interpret this field as stop on consecutive setmarks.

A recover object buffer order (ROBO) bit of one specifies logical blocks shall be returned from the object buffer of the logical unit on a RECOVER BUFFERED DATA command in LIFO order (last-in-first-out) from that they were written to the object buffer. A RBO bit of zero specifies logical blocks shall be returned in FIFO (first-in-first-out) order.

A report early-warning (REW) bit of zero specifies the device server shall not report the early-warning condition for read operations and it shall report early-warning at or before any medium-defined early-warning position during write operations.

A REW bit of one specifies the device server shall return CHECK CONDITION status. The additional sense code shall be set to END-OF-PARTITION/MEDIUM DETECTED, and the EOM bit set to one in the sense data when the early-warning position is encountered during read and write operations. If the REW bit is one and the SEW bit is zero, the device server shall return CHECK CONDITION status with the sense key set to VOLUME OVERFLOW when early-warning is encountered during write operations.

NOTE 43 A REW bit of one is intended for compatibility with application clients using old tape formats that require an early-warning indication during read operations. Other application clients should set this bit to zero to avoid potential data loss when interchanging tapes between devices.

The GAP SIZE field value determines the size of the inter-block gap when writing data. A value of 00h specifies the device's defined gap size. A value of 01h specifies a device defined gap size sufficiently long to support update-in-place. Values of 02h through 0Fh are multipliers on the device's defined gap size. Values 10h through 7Fh are reserved. Values 80h through FFh are vendor-specific.

The EOD DEFINED field specifies the format type that the logical unit shall use to detect and generate the EOD area. The values for EOD DEFINED are specified in table 61.

Table 61 — EOD DEFINED values

Code	Description
000b	Logical unit’s default EOD definition
001b	Format-defined erased area of medium
010b	As specified in the SOCF field
011b	EOD recognition and generation is not supported
100b - 111b	Reserved

An enable EOD generation (EEG) bit set to one specifies the logical unit shall generate the appropriate EOD area, as determined by the EOD field. A value of zero specifies EOD generation is disabled.

NOTE 44 Some logical units may not generate EOD at the completion of any write-type operation.

A synchronize at early-warning (SEW) bit set to one specifies the logical unit shall cause any buffered logical objects to be transferred to the medium prior to returning status when positioned between early-warning and EOP. A SEW bit of zero specifies the logical unit may retain unwritten buffered logical objects in the object buffer when positioned between early-warning and EOP (see 5.6, 5.7, 6.8, and 6.9).

A software write protection (SWP) bit of one specifies the device server shall perform a synchronize operation then enter the write-protected state (see 4.2.12 and 4.2.12.3). A SWP bit of zero specifies the device server

A software write protection (SWP) bit of one specifies the logical unit shall inhibit all writing to the medium after writing all buffered data, if any (see 4.2.12 and 4.2.12.3). When the SWP bit is one, all commands requiring eventual writes to the medium shall return CHECK CONDITION status. The sense key shall be set to DATA PROTECT and the additional sense code shall be set to LOGICAL UNIT SOFTWARE WRITE PROTECTED. A SWP bit of zero specifies the logical unit may inhibit writing to the medium, dependent on other write inhibits.

A block address mode lock (BAML) bit of zero specifies the selection of the block address mode shall be determined based on the first block address mode unique command that is received after a successful load operation or a successful completion of a command that positions the medium to BOP. A BAML bit of one specifies the selection of the block address mode shall be determined based on the setting of the BAM bit. See 4.2.15 for a description of block address mode selection.

The block address mode (BAM) bit is valid only if the BAML bit is set to one. If the BAML bit is set to zero, the BAM bit shall be ignored. If the BAML bit is set to one and the BAM bit is set to zero, the logical unit shall operate using implicit address mode. If the BAML bit is set to one and the BAM bit is set to one, the logical unit shall operate using explicit address mode. See 4.2.15 for a description of block address mode selection.

The OBJECT BUFFER SIZE AT EARLY WARNING field specifies the value, in bytes, that the logical unit shall reduce its logical object buffer size to when writing in a position between its early-warning and end-of-partition. A value of zero specifies the implementation of this function is vendor-specific.

> NOTE 45 The intent is to prevent the loss of data by limiting the size of the object buffer when near the end-of-partition.

The SELECT DATA COMPRESSION ALGORITHM field set to 00h specifies the logical unit shall not use a compression algorithm on any data sent to it prior to writing the data to the medium. A value of 01h specifies the data to be written shall be compressed using the logical unit's default compression algorithm. Values 02h through 7Fh are reserved. Values 80h through FFh are vendor-specific. The SELECT DATA COMPRESSION ALGORITHM field shall be ignored if a Data Compression mode page with the DCE bit set to one is received by the device in the same MODE SELECT command.

> NOTE 46 New implementations use the Data Compression mode page (see 8.3.2) for specifying data compression behavior.

An only if reserved (OIR) bit set to one specifies that the device server shall only perform a command if a reservation or persistent reservation exists which allows access to the I_T nexus from which the command was received. When OIR is one and a command is received from an I_T nexus for which no reservation exists, the device server shall not perform the command. When OIR is one and a command is received from an I_T nexus for a logical unit or element upon which no reservation or persistent reservation exists, the device server shall terminate the command with CHECK CONDITION status, and shall set the sense key to NOT READY and the additional sense code and qualifier to NOT RESERVED. Commands that shall not be effected by OIR are RESERVE, RELEASE, PERSISTENT RESERVE IN AND PERSISTENT RESERVE OUT. Commands affected by OIR are defined with reference to the tables that define the commands allowed in the presence of various reservations in this standard (Table 10) and in the SPC-2 & SPC-3 standards. Any command which has “Conflict” in any column of those tables shall be effected by OIR, except those noted above as not being effected.

The REWIND ON RESET field is specified in table 62. The REWIND ON RESET field, if implemented, shall be persistent across logical unit resets.

Table 62 — REWIND ON RESET field definition

Code	Description
00b	Vendor-specific
01b	The logical unit shall position to the beginning of the default data partition (BOP 0) on logical unit reset.
10b	The logical unit shall maintain its position on logical unit reset.
11b	Reserved

An associated write protection (ASOCWP) bit of one specifies the logical unit shall inhibit all writing to the medium after performing a synchronize operation (see 4.2.12 and 4.2.12.4). When the ASOCWP bit is one, the currently mounted volume is logically write protected until the volume is de-mounted (see 4.2.12 and 4.2.12.4). When the ASOCWP bit is one, all commands requiring eventual writes to the medium shall return CHECK CONDITION status. The sense key shall be set to DATA PROTECT and the additional sense code shall be set to ASSOCIATED WRITE PROTECT. An ASOCWP bit of zero specifies the currently mounted volume is not write protected by the associated write protection. The ASOCWP bit shall be set to zero by the device server when the volume is de-mounted. This change of state shall not cause a unit attention condition. If the application client sets the ASOCWP bit to one while no volume is mounted, the device server shall terminate the MODE SELECT command with CHECK CONDITION status. The sense key shall be set to NOT READY and the additional sense code shall be set to MEDIUM NOT PRESENT. If the Device Configuration mode page is savable, the ASOCWP bit shall be saved as zero, regardless of the current setting.

A persistent write protection (PERSWP) bit of one specifies the currently mounted volume is logically write protected (see 4.2.12 and 4.2.12.5). When the PERSWP bit is one, all commands requiring eventual writes to the medium shall return CHECK CONDITION status. The sense key shall be set to DATA PROTECT and the additional sense code shall be set to PERSISTENT WRITE PROTECT. A PERSWP bit of zero specifies the currently mounted volume is not write protected by the persistent write protection. The PERSWP bit shall be set to zero by the device server when the volume is de-mounted or when a volume is mounted with persistent write protection disabled. The PERSWP shall be set to one by the device server when a volume is mounted with persistent write protection enabled. These changes of state shall not cause a unit attention condition. If the application client sets the PERSWP bit to one while no volume is mounted, the device server shall terminate the MODE SELECT command with CHECK CONDITION status. The sense key shall be set to NOT READY. The additional sense information shall be set to MEDIUM NOT PRESENT. If the application client sets the PERSWP bit to one when the logical position is not at BOP 0, the device server shall return CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST. The additional sense information shall be set to POSITION PAST BEGINNING OF MEDIUM. If the Device Configuration mode page is savable, the PERSWP bit shall be saved as zero, regardless of the current setting.

A permanent write protection (PRMWP) bit of one specifies the currently mounted volume is logically write protected (see 4.2.12 and 4.2.12.6). When the PRMWP bit is one, all commands requiring eventual writes to the medium shall return CHECK CONDITION status and the sense key shall be set to DATA PROTECT and the additional sense code shall be set to PERMANENT WRITE PROTECT. A PRMWP bit of zero specifies the currently mounted volume is not write protected by the permanent write protection. The PRMWP bit shall be set to zero by the device server when the volume is de-mounted or when a volume is mounted with permanent write protection disabled. The PRMWP shall be set to one by the device server when a volume is mounted with permanent write protection enabled. These changes of state shall not cause a unit attention condition. If the application client sets the PRMWP bit to one while no volume is mounted, the device server shall terminate the MODE SELECT command with CHECK CONDITION status. The sense key shall be set to NOT READY. The additional sense information shall be set to MEDIUM NOT PRESENT. If the application client sets the PRMWP bit to one when the logical position is not at BOP 0, the device server shall return CHECK CONDITION status. The sense key shall be set to ILLEGAL REQUEST. The additional sense information shall be set to POSITION PAST BEGINNING OF MEDIUM. If the application client attempts to change the PRMWP bit from one to zero, the device server shall terminate the MODE SELECT command with CHECK CONDITION status. The sense key shall be set to DATA PROTECT. The additional sense information shall be set to PERMANENT WRITE PROTECT. If the Device Configuration mode page is savable, the PRMWP bit shall be saved as zero, regardless of the current setting.

SPC-3

Assign an additional ASC/ASCQ to indicate NOT RESERVED.